Using a Reaction-Based Differential-Algebraic Equation Approach for Reactive Transport Modeling in Hydrologic Systems*

Hwai-Ping Cheng† Stacy E. Howington†

Abstract

We discuss how a Reaction-Based, Differential-Algebraic Equation (RB-DAE) approach can provide robust numerical solutions for biogeochemical systems that are composed of fast, slow, reversible, and irreversible reactions. We also discuss several numerical strategies for solving reactive transport equations using high performance computing (HPC).

1 Problem Specification

Computer simulation of flow and transport is an essential component in the rigorous analysis of water supply, contamination, environmental cleanup, and ecosystem restoration. Simulation of coupled equations for transport and biogeochemical reactions based on the principle of conservation is the only quantitative approach to date for integrating multiple, complex, environmental processes into an internally consistent conceptual model with which to assess water quality and to design engineered solutions for remedial alternatives. A reactive transport (RT) model in a more general sense treats a multi-component, multi-species system in which a number of equilibrium-controlled (fast reversible) and perhaps kinetic (slow) and instantaneous (fast irreversible) reactions occur simultaneously. In a typical hydrologic cycle, water moves on, above, or below the surface of the earth, and its speed can vary over many orders of magnitude. Similarly, the rate of a biogeochemical reaction, whether natural or man-induced, can change drastically over time and space. The combination of these two facts makes RT modeling in hydrologic systems an extremely difficult task.

1.1 RT Equations in Primitive Form. A typical set of RT equations include transport processes (e.g., advection, diffusion, dispersion), biogeochemical reactions, and sources/sinks. They can be written in the so-called primitive form as

$$\frac{\partial \mathbf{C}}{\partial t} = L(\mathbf{C}) + \mathbf{SS_R} + \mathbf{SS_C}, \tag{1.1}$$

where $\mathbf{C}$ is the species concentration vector; $L()$ denotes the linear transport operator that accounts for advection, diffusion, and dispersion; $\mathbf{SS_R}$ represents nonlinear sources/sinks due to biogeochemical reactions; and $\mathbf{SS_C}$ represents linear sources/sinks due to other activities, such as injection or extraction. For immobile species, the transport and source/sink terms in Eq. (1.1) may be neglected.

As previously mentioned, $L(\mathbf{C})$, $\mathbf{SS_R}$, and $\mathbf{SS_C}$ in Eq. (1.1) may vary over wide ranges. While $L(\mathbf{C})$ and $\mathbf{SS_C}$ fall in specific ranges as defined by the associated hydrologic system, the range of $\mathbf{SS_R}$ depends on the characteristics of the reactions taken into account. The relative importance of reaction and transport at a specific distance scale L can be described by the non-dimensional *DamKöhler* number [1], *Da*, which is defined as

$$Da = \frac{\tau_{trans}}{\tau_{rxn}}, \tag{1.2}$$

where τ_{trans} is the transport characteristic time required for water to traverse the specific distance L and τ_{rxn} is the reaction characteristic time required for a reaction to reach equilibrium [2]. Each reaction, if reversible, has its own τ_{rxn}. The values of *Da* can change drastically over space and time in an RT system, which implies both transport processes and biogeochemical reactions must be resolved adequately in space and time so that Eq. (1.1) can be solved accurately.

1.2 Reaction Model. A biogeochemical reaction model can be represented in the primitive form as

$$\frac{d\mathbf{C}}{dt} = \mathbf{SS_R}. \tag{1.3}$$

No matter which solution technique is used in solving RT equations, it is vital to model biogeochemical reactions and solve for concentration distributions among species accurately and efficiently. Bethke [2] discussed the conceptual model, mathematical formulations, and numerical solutions for reaction processes and systems of various kinds. His discussion demonstrates the capabilities of a reaction model and determines the model's effect on the performance of reactive transport modeling, which includes solute transport. For this reason, a powerful biogeochemical reaction model should be general enough to account for a wide spectrum of reactions. Some reactions are slow and others are fast

* Supported by the System-Wide Water Resources Program (SWWRP) of U.S. Army Corps of Engineers; Allowed to present by the chief of US Army Corps of Engineers.

†Coastal and Hydraulics Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS.

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

relative to the temporal components of the model [3], (e.g., modeling time step and the residence time within a model element or cell). Therefore, the reaction model must approximate systems that are combinations of equilibrium-controlled, instantaneous, and kinetic reactions.

1.3 Challenges in Solving RT Equations. The RT equations must be approximated in a way that maintains non-negative concentrations, preserves mass conservation, maintains steep gradients, and provides answers in a reasonable simulation time. Non-negativity of concentration is essential in RT computation [2]. Negative concentrations caused by inadequate numerical solutions will migrate through transport computation and may cause convergence issues or nonrealistic results in the reaction computation. Mass conservation is another requirement in RT computation. Mass must be conserved both in the solute transport and reaction computations at all times. Treatment limits for many contaminants are very small. For example, the maximum contamination level set by the US Environmental Protection Agency for both tetrachloroethylene (PCE) and trichloroethylene (TCE) is 5.0 ppb, while their solubility limits are $1.5\text{x}10^5$ and $1.28\text{x}10^6$ ppb, respectively. Compared with initial plume concentrations, even small errors in mass conservation may lead to invalid conclusions about the success of remedial measures. Sharp fronts are common in RT due to reactions [4], boundaries, and sources or sinks. Excessive numerical spreading when approximating these fronts [4] may produce misleading simulation results. Fang et al. explained seven drawbacks associated with solving the primitive form of reaction equations [5]. One drawback concerning computational efficiency is that time step size must be sufficiently small to resolve the fastest reaction when solving the primitive-form reaction model. Otherwise, incorrect solutions or convergence issues will appear. As a consequence, the primitive RT equations can require tremendous computational resources when using small time steps to resolve fast reactions. In hydrogeological systems, where most reactions are equilibrium-controlled, or when instantaneous reactions are introduced for cleanup purposes, it may be impractical, even with high performance computing, to solve the primitive RT equations for multi-year simulations because extremely small time steps would be needed to resolve very fast reactions.

To summarize, negative concentrations, mass conservation errors, and non-physical spreading must be avoided for robust and accurate RT computation; efficient RT computation cannot be achieved in solving the primitive RT equations when equilibrium-controlled and/or instantaneous reactions exist. In the following sections, we first discuss how a reaction-based differential-algebraic equation (RB-DAE) approach can overcome the afore-mentioned time-step drawback associated with the pri-mitive RT equations. Then we discuss the pros and cons of some solution techniques related to solving the RT equations in HPC when the RB-DAE approach is employed. Although it is beyond the scope of this paper, solving advection-diffusion-dispersion transport equations is a prerequisite to RT computation. There are many research papers discussing the pros and cons of various solution techniques for solving transport equations. Yeh provided a comprehensive discussion on several main-stream numerical methods to solve advection-dominant transport equations [4].

2 Primitive Approach vs. RB-DAE Approach

2.1 Reaction Equations. The DAE approach [5, 6] models equilibrium-controlled reactions with mass action equations to overcome the time step drawback associated with the primitive approach. Fang et al. [7, 8] proposed a general paradigm to generate and solve reaction equations automatically and systematically, where an RB-DAE approach was presented. In their RB-DAE approach, the governing equations consist of reaction-based rate equations and mass conservation equations, which are mathematically equivalent to the species-based rate equations used in the primitive approach. We extended their RB-DAE approach to construct the governing equations for multi-species, multi-reaction biogeochemical modeling systems [9], which include

- one equilibrium equation for each independent equilibrium-controlled reaction,
- one constraint equation for each independent instantaneous reaction,
- one rate equation for each independent kinetic reaction,
- one total mass conservation equation for each basis (component) species,

where equilibrium and constraint equations are algebraic equations, the rate and mass conservation equations are ordinary differential equations, and the number of basis species is equal to the total number of species minus the number of independent reactions.

To demonstrate this extended RB-DAE approach, we consider a hypothetical reaction system that contains 12 species and five independent reactions. The 12 species are labeled A, B, C, D, E, F, G, H, M, N, P, and Q. Among the five reactions, (R3) and (R4) are equilibrium-controlled reactions, (R5) is an instantaneous reaction, and (R1) and (R2) are kinetic reactions.

(R1) $A + 2B \rightarrow C$ (kinetic)

(R2) $D + 2E \leftrightarrow F$ (kinetic)

(R3) $C + 2F \leftrightarrow 2G + H$ (equilibrium-controlled)

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

(R4) $\mathrm{H} + 2\mathrm{M} \leftrightarrow \mathrm{N}$ (equilibrium-controlled)

(R5) $\mathrm{N} + 2\mathrm{P} \rightarrow \mathrm{Q}$ (instantaneous)

The species-based rate equations for the governing equation using the primitive approach are

$$\frac{d[A]}{dt} = -R_1, \tag{2.1}$$

$$\frac{d[B]}{dt} = -2R_1, \tag{2.2}$$

$$\frac{d[C]}{dt} = R_1 - R_3, \tag{2.3}$$

$$\frac{d[D]}{dt} = -R_2, \tag{2.4}$$

$$\frac{d[E]}{dt} = -2R_2, \tag{2.5}$$

$$\frac{d[F]}{dt} = R_2 - 2R_3, \tag{2.6}$$

$$\frac{d[G]}{dt} = 2R_3, \tag{2.7}$$

$$\frac{d[H]}{dt} = R_3 - R_4, \tag{2.8}$$

$$\frac{d[M]}{dt} = -2R_4, \tag{2.9}$$

$$\frac{d[N]}{dt} = R_4 - R_5, \tag{2.10}$$

$$\frac{d[P]}{dt} = -2R_5, \tag{2.11}$$

and

$$\frac{d[Q]}{dt} = R_5, \tag{2.12}$$

where [ ] denotes concentration and R_1 through R_5 are reaction rates of (R1) through (R5), respectively. In general, R_1 through R_5 are functions of the concentrations of species which are associated with the corresponding reactions. Eqs. (2.1) through (2.12) pose a set of nonlinear ordinary differential equations.

The reaction equations constructed with the RB-DAE approach can be obtained by preprocessing with matrix decomposition. We briefly describe this preprocessing here, while the details of the preprocessing can be found elsewhere [9].

Eqs. (2.1) through (2.12) can be written in matrix form as

$$\mathbf{I}\frac{d\mathbf{G}}{dt} = \mathbf{v}\mathbf{R}, \tag{2.13}$$

where $\mathbf{I}$ is the identity matrix of size 12; $\mathbf{G}$ is the concentration vector representing the 12 species; $\mathbf{R}$ is the vector representing the 5 reaction rates, *i.e.*, $\{R_1, R_2, R_3, R_4, R_5\}^{\mathrm{T}}$; and $\mathbf{v}$ is the reaction stoichiometry matrix, which can be written as

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In the preprocessing, we decompose Eq. (2.13) to generate reaction-based rate equations corresponding to the five independent reactions and mass conservation equations associated with seven (= 12 - 5) basis species. To achieve this, we first pick basis species, followed by implementing Gauss-Jordan elimination with full pivoting to decompose the reaction network. The selection of basis species can be arbitrary [7]. Different selections will produce different decomposition results, but all the decomposition results are mathematically equivalent. Such decomposition will generate the following matrix equation.

$$\begin{bmatrix} \mathbf{A_1} \\ \mathbf{A_2} \end{bmatrix} \frac{d\mathbf{G}}{dt} = \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \end{bmatrix} \mathbf{R_1}, \tag{2.14}$$

where $\mathbf{A_1}$ is a 5x12 coefficient matrix; $\mathbf{A_2}$ is a 7x12 coefficient matrix, $\mathbf{D}$ is a 5x5 coefficient matrix; and $\mathbf{0}$ is a 7x5 zero-coefficient matrix.

Fang et al. [7] have described detailed step-by-step decomposition procedures, and we do not repeat the lengthy description here. Although they did not include instantaneous reactions, the essential details concerning decomposition were clearly explained in their paper.

When we pick species A, B, D, E, H, N, and P as basis species, the coefficient matrices in Eq. (2.15) are

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

$$\mathbf{A_1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{A_2} = \begin{bmatrix} 2 & 0 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix},$$

and

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Eq. (2.14) represents a set of ordinary differential equations comprising five reaction-based rate equations and seven mass conservation equations, as given below.

$$\frac{d[G]}{dt} = -R_3, \tag{2.15}$$

$$\frac{d[M]}{dt} = -2R_4, \tag{2.16}$$

$$\frac{d[Q]}{dt} = R_5, \tag{2.17}$$

$$-2\frac{d[C]}{dt} - \frac{d[G]}{dt} = -2R_1, \tag{2.18}$$

$$\frac{d[F]}{dt} + \frac{d[G]}{dt} = R_2, \tag{2.19}$$

$$2\frac{d[A]}{dt} + 2\frac{d[C]}{dt} + \frac{d[G]}{dt} = 0, \tag{2.20}$$

$$\frac{d[B]}{dt} + 2\frac{d[C]}{dt} + \frac{d[G]}{dt} = 0, \tag{2.21}$$

$$\frac{d[D]}{dt} + \frac{d[F]}{dt} + \frac{d[G]}{dt} = 0, \tag{2.22}$$

$$\frac{d[E]}{dt} + 2\frac{d[F]}{dt} + 2\frac{d[G]}{dt} = 0, \tag{2.23}$$

$$\frac{d[G]}{dt} - 2\frac{d[H]}{dt} + \frac{d[M]}{dt} = 0, \tag{2.24}$$

$$\frac{d[M]}{dt} + 2\frac{d[N]}{dt} + 2\frac{d[Q]}{dt} = 0, \tag{2.25}$$

and

$$\frac{d[P]}{dt} + 2\frac{d[Q]}{dt} = 0, \tag{2.26}$$

where Eqs. (2.15) through (2.19) are the reaction-based rate equations corresponding to (R3), (R4), (R5), (R1), and (R2), respectively, and Eqs. (2.20) through (2.26) are the total mass conservation equations of basis species A, B, D, E, H, N, and Q, respectively. To avoid using extremely small time steps in biogeochemistry computation, equilibrium-controlled and instantaneous reactions are represented by equilibrium and constraint equations, respectively, rather than by rate equations. Eqs. (2.15) through (2.17) are thus replaced by

$$K_3^{eq} = f([G],[H],[C],[F]), \tag{2.27}$$

$$K_4^{eq} = g([N],[H],[M]), \tag{2.28}$$

and

$$[N] = 0 \quad \text{or} \quad [P] = 0, \tag{2.29}$$

where f and g are functions used to describe equilibrium for (R3) and (R4), respectively; K_3^{eq} and K_4^{eq} are the corresponding equilibrium constants; and Eq. (2.29) is the constraint equation for (R5). In this case, nonlinearity appears in only four equations: (2.18), (2.19), (2.27), and (2.28), rather than all the 12 reaction equations. This suggests the RB-DAE equations may produce faster convergence than the primitive equations, even though they are mathematically equivalent.

Algebraic equations (2.27) through (2.29) are employed to represent fast (reversible and irreversible) reactions, and R_3 through R_5 do not appear in the reaction equations when the RB-DAE approach is employed. As a result, time steps need to be small enough only to resolve slow reactions (R1) and (R2), rather than all fast and slow reactions.

Figure 1 compares the computed concentration profiles of species N, P, and Q when four different time steps (*i.e.*, dt = 10, 1, 0.1, and 0.01) were used in the primitive approach (top) and in the RB-DAE approach (bottom), where the Newton-Raphson method was used to solve the reaction equations. It shows that when the intuitive scheme [10] was used to deal with instantaneous reaction in the primitive approach, the numerical solutions were reasonable when very small time steps (*i.e.*, dt = 0.1 and 0.01) were used but were erroneous when larger time steps were used (*i.e.*, dt = 10 and 1). On the other hand, the RB-DAE approach provides

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

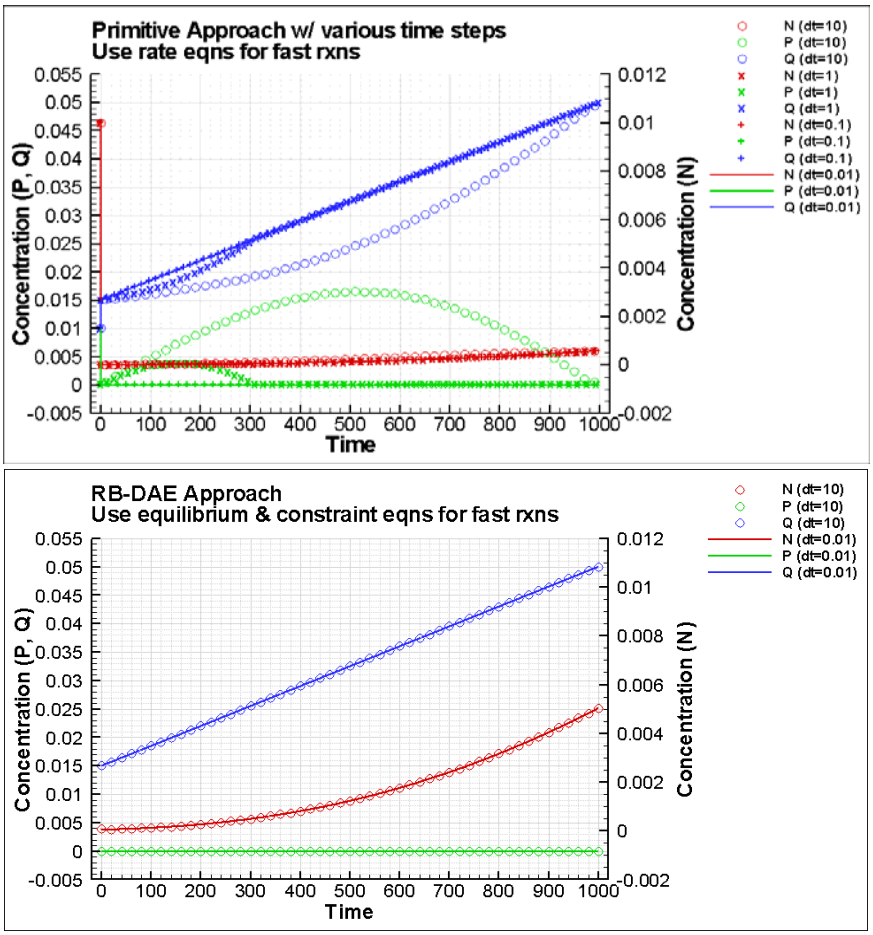

Figure 1: Comparison of the computed concentrations from the primitive and the RB-DAE approaches.

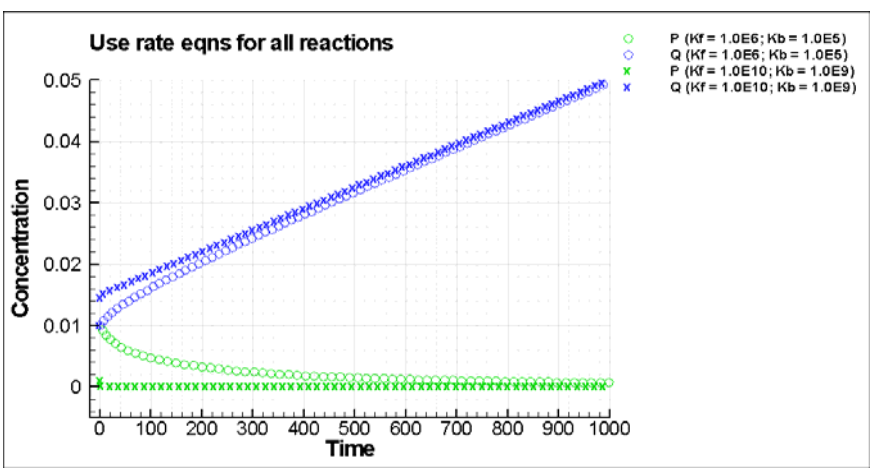

Figure 2: Comparison of the computed concentrations when different rate constants are used to represent fast reactions.

accurate solutions even a large time step (*i.e.*, dt = 10) was used.

The forward and backward rate constants were set to 10^{10} and 10^{9}, respectively, to mimic fast reactions in the

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

primitive approach to obtain numerical results in Figure 1. When the rate constants were set to 10^6 and 10^5 (*i.e.,* smaller values) to mimic fast reactions, however, significant differences on the concentration profiles of species P and Q exist (Figure 2), indicated that the simulation result was inaccurate if reaction rate constants were not large enough to represent fast reactions using the primitive approach.

Through this example, we demonstrate (1) reaction rate constants need to be set large enough to adequately mimic fast reactions in the primitive approach, and (2) the RB-DAE approach provides accurate solutions even when using large time steps for fast reactions.

2.2 RT Equations. If there are n mobile species in the aforementioned hypothetical reaction system, there are n transport-reaction equations and (12-n) rate equations to solve simultaneously when the primitive approach is used. They can be written as given in Eq. (1.1) where $\mathbf{SS_R} = (-R_1, -2R_1, R_1-R_3, -R_2, -2R_2, R_2-2R_3, R_3, R_3-R_4, -R_4, R_4-R_5, -2R_5, R_5)$.

Suppose species M, N, and Q are immobile species (*i.e.*, not subject to transport), the RT equations associated with the RB-DAE approach are

$$K_3^{eq} = f([G],[H],[C],[F]), \tag{2.30}$$

$$K_4^{eq} = g([N],[H],[M]), \tag{2.31}$$

$$[N] = 0 \quad \text{or} \quad [P] = 0, \tag{2.32}$$

$$-2\frac{d[C]}{dt} - \frac{d[G]}{dt} + L(-2[C]-[G]) = SS_{-2[C]-[G]} - 2R_1, \tag{2.33}$$

$$\frac{d[F]}{dt} + \frac{d[G]}{dt} + L([F]+[G]) = SS_{[F]+[G]} + R_2, \tag{2.34}$$

$$2\frac{d[A]}{dt} + 2\frac{d[C]}{dt} + \frac{d[G]}{dt} + L(2[A]+2[C]+[G]) = SS_{2[A]+2[C]+[G]}, \tag{2.35}$$

$$\frac{d[B]}{dt} + 2\frac{d[C]}{dt} + \frac{d[G]}{dt} + L([B]+2[C]+[G]) = SS_{[B]+2[C]+[G]}, \tag{2.36}$$

$$\frac{d[D]}{dt} + \frac{d[F]}{dt} + \frac{d[G]}{dt} + L([D]+[F]+[G]) = SS_{[D]+[F]+[G]}, \tag{2.37}$$

$$\frac{d[E]}{dt} + 2\frac{d[F]}{dt} + 2\frac{d[G]}{dt} + L([E]+2[F]+2[G]) = SS_{[E]+2[F]+2[G]}, \tag{2.38}$$

$$\frac{d[G]}{dt} - 2\frac{d[H]}{dt} + \frac{d[M]}{dt} + L([G]-2[H]) = SS_{[G]-2[H]}, \tag{2.39}$$

$$\frac{d[M]}{dt} + 2\frac{d[N]}{dt} + 2\frac{d[Q]}{dt} = 0, \tag{2.40}$$

and

$$\frac{d[P]}{dt} + 2\frac{d[Q]}{dt} + L([P]) = SS_{[P]}, \tag{2.41}$$

where $L()$ is the transport operator; SS is the prescribed source/sink term; the subscript SS defines the composition of the source/sink term in each of Eqs. (2.33) through (2.41).

Among these 12 RT equations associated with the RB-DAE approach, Eqs. (2.30) through (2.32) characterize the three fast reactions: (R3), (R4), and (R5); Eqs. (2.33) and (2.34) describe the transport of kinetic variables [11] associated with the two slow reactions: (R1) and (R2), where the kinetic variable is (-2[C]-[G]) in Eq. (2.33) and ([F]+[G]) in Eq. (2.34); Eqs. (2.35) through (2.38) are transport equations for the total mass of basis species A, B, D, and E, respectively, where the total mass remains in the mobile phase; Eqs. (2.39) and (2.41) account for the conservation of the total mass of basis species H and P, respectively, where only the dissolved mass is transportable (*i.e.*, retardation exists); and Eq. (2.40) is the conservation of the total mass of basis species N, where all the associated species (i.e., M, N, and Q) are immobile. As a result, the 12 RT equations include two nonlinear equilibrium equations: Eqs. (2.30) and (2.31), one linear constraint equation: Eq. (2.32), four nonlinear transport equations: Eqs. (2.33), (2.34), (2.39), and (2.41), four linear transport equations: Eqs. (2.35) through (2.38), and one conservation equation: Eq. (2.40). It is noted that the number of nonlinear transport equations with the RB-DAE approach is four, which is fewer than nine with the primitive approach (the three immobile species are not subject to transport).

Our discussion in this section indicates that the RB-DAE approach is superior to the primitive approach in RT computation because (1) the RB-DAE approach does not need extremely small time steps to resolve fast reactions but the primitive approach does, and (2) the RB-DAE approach contains fewer nonlinear reaction equations as well as fewer nonlinear transport equations.

3 RT Solution Techniques.

The RT equations in the RB-DAE approach can be written as the combination of

$$\Phi(\mathbf{C}) = 0, \tag{3.1}$$

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

$$\frac{\partial \mathbf{K}}{\partial t} = L(\mathbf{K_m}) + \mathbf{SS_{R,K}} + \mathbf{SS_{C,K_m}}, \tag{3.2}$$

and

$$\frac{\partial \mathbf{T}}{\partial t} = L(\mathbf{T_m}) + \mathbf{SS_{C,T_m}}, \tag{3.3}$$

where Eq. (3.1) represents algebraic equations used to describe equilibrium-controlled and instantaneous reactions; Eq. (3.2) states the reactive transport of kinetic variables (**K**) where only the mobile part (**K_m**) is transportable; and Eq. (3.3) describes the transport of the total mass of basis species (**T**) where the mobile portion (**T_m**) is subject to transport.

When a kinetic variable contains only mobile species, we have $K_m = K$ for that variable. When the total mass of a basis stays in the mobile phase, we have $T_m = T$ for that total mass. When a total mass involves only the immobile species, we have $L(T_m) = 0$ and $SS_{R,Tm} = 0$. Based on this, Eq. (3.2) can be further written as the combination of

$$\frac{\partial(\mathbf{K1})}{\partial t} = L(\mathbf{K1}) + \mathbf{SS_{R,K1}} + \mathbf{SS_{C,K1}}, \tag{3.4}$$

and

$$\frac{\partial(\mathbf{K2})}{\partial t} = L(\mathbf{K2_m}) + \mathbf{SS_{R,K2}} + \mathbf{SS_{C,K2_m}}, \tag{3.5}$$

where **K1** denotes all kinetic variables that contain mobile species only; **K2** denotes all kinetic variables that contains both mobile and immobile species. Likewise, Eq. (3.3) can be written as the combination of

$$\frac{\partial(\mathbf{T1})}{\partial t} = L(\mathbf{T1}) + \mathbf{SS_{C,T1}}, \tag{3.6}$$

$$\frac{\partial(\mathbf{T2})}{\partial t} = L(\mathbf{T2_m}) + \mathbf{SS_{C,T2_m}}, \tag{3.7}$$

and

$$\frac{\partial(\mathbf{T3})}{\partial t} = 0, \tag{3.8}$$

where **T1** denotes all total masses that stay in the mobile phase; **T2** denotes all total masses that involves both mobile and immobile species; **T3** denotes all total masses that are associated with immobile species only.

For research applications, accuracy is essential for helping to study processes and design reaction systems. For large field-scale problems, on the other hand, efficiency and robustness are needed to maintain accuracy within the bounds of uncertainty associated with model parameters. Although high performance computing can reduce run time considerably, it is still essential to use adequate numerical strategies to achieve modeling purposes.

3.1 SNIA, SIA, and DSA. The sequential non-iterative approach (SNIA), the sequential iterative approach (SIA), and direct substituting approach (DSA) are three approaches commonly used for solving RT equations [4]. The SNIA solves first for transport for each mobile species then for biogeochemistry independently at each node or cell at each time step [12]. Explicit SNIA is easy to implement for high performance computing but requires small time steps to avoid stability problems. SIA executes nonlinear iterations between transport and biogeochemistry computation until convergence at each time step to obtain accurate results [4, 12]. In the DSA, the equilibrium equations are substituted into the reactive transport equations to solve N_p nonlinear partial differential equations simultaneously. N_p is the number of primary dependent variables (PDV) that is equal to the total number of species minus the number of secondary species that can be represented by basis species through the equilibrium equations [4]. In serial computation, DSA is not practical for multi-dimensional simulation because it requires a simultaneous solution of a significant number of field equations, which demands excessive CPU memory and time. DSA may be affordable in high performance computing if the number of PDV is small. In this regard, iterative approaches are still more practical than the DSA in high performance RT computing.

In the following, we focus on three sequential approaches for the RT system composed of Eqs. (3.4) through (3.8). These three approaches are operator splitting (OS), predictor-corrector (PC), and complete iteration (CI), where OS and PC are SNIA schemes, and CI is an SIA approach.

3.2 Operator Splitting. With OS, the discretized equations solved in the transport step include

$$\frac{\mathbf{K1}^{OS} - \mathbf{K1}^{(n)}}{\Delta t} = L(\mathbf{K1}^{OS}) + \mathbf{SS}^{(n+1)}_{\mathbf{C,K1}}, \tag{3.9}$$

$$\frac{\mathbf{K2}^{OS} - \mathbf{K2}^{(n)}}{\Delta t} = L(\mathbf{K2}^{OS} - \mathbf{K2}^{(n)}_{\mathbf{im}}) + \mathbf{SS}^{(n+1)}_{\mathbf{C,K2_m}}, \tag{3.10}$$

$$\frac{\mathbf{T1}^{(n+1)} - \mathbf{T1}^{(n)}}{\Delta t} = L(\mathbf{T1}^{(n+1)}) + \mathbf{SS}^{(n+1)}_{\mathbf{C,T1}}, \tag{3.11}$$

and

$$\frac{\mathbf{T2}^{(n+1)} - \mathbf{T2}^{(n)}}{\Delta t} = L(\mathbf{T2}^{(n+1)} - \mathbf{T2}^{(n)}_{\mathbf{im}}) + \mathbf{SS}^{(n+1)}_{\mathbf{C,T2_m}}, \tag{3.12}$$

where Δt is time interval; $\mathbf{K2}^{(n)}_{\mathbf{im}}$ and $\mathbf{T2}^{(n)}_{\mathbf{im}}$ are the immobile part of **K2** and **T2**, respectively, at the old time

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

step; superscript *OS* is used to represent the intermediate concentration; and superscripts (n) and ($n+1$) are used to represent the old and the new time steps, respectively. The solutions of Eqs. (3.9) through (3.12) are then taken to solve the following three equations plus the conservation equations of the total mass of basis species in the reaction step.

$$\Phi\left(\mathbf{C}^{(n+1)}\right) = 0, \tag{3.13}$$

$$\frac{\mathbf{K1}^{(n+1)} - \mathbf{K1}^{OS}}{\Delta t} = \mathbf{SS}_{\mathbf{R,K1}}^{(n+1)}, \tag{3.14}$$

and

$$\frac{\mathbf{K2}^{(n+1)} - \mathbf{K2}^{OS}}{\Delta t} = \mathbf{SS}_{\mathbf{R,K2}}^{(n+1)}. \tag{3.15}$$

The equations included in the reaction step are solved simultaneously on a node-by-node or cell-by-cell basis, which is easily parallelized for high performance computing. The full-pivoting direct solver is probably the best way to ensure numerical stability in solving biogeochemistry, where species concentration can range over many orders of magnitude. The computed species concentrations at the current time step are used to update $\mathbf{K2}_{\mathbf{im}}^{(n)}$ and $\mathbf{T2}_{\mathbf{im}}^{(n)}$, which will be used in the transport step of the next time interval.

3.3 Predictor-Corrector. The PC technique differs from the OS technique in that the source/sink terms due to reactions from the old time step (*i.e.*, $\mathbf{SS}_{\mathbf{R,K1}}^{(n)}$ and $\mathbf{SS}_{\mathbf{R,K2}}^{(n)}$) are used as the predicted reaction contribution in solving for **K1** and **K2**. Therefore, the equations included in the transport step are Eqs. (3.11), (3.12) and

$$\frac{\mathbf{K1}^{PC} - \mathbf{K1}^{(n)}}{\Delta t} = L\left(\mathbf{K1}^{PC}\right) + \mathbf{SS}_{\mathbf{C,K1}}^{(n+1)} + \mathbf{SS}_{\mathbf{R,K1}}^{(n)}, \tag{3.16}$$

and

$$\frac{\mathbf{K2}^{PC} - \mathbf{K2}^{(n)}}{\Delta t} = L\left(\mathbf{K2}^{PC} - \mathbf{K2}_{\mathbf{im}}^{(n)}\right) + \mathbf{SS}_{\mathbf{C,K2_m}}^{(n+1)} + \mathbf{SS}_{\mathbf{R,K1}}^{(n)}, \tag{3.17}$$

where superscript *PC* is used to represent the intermediate concentration. The equations solved in the reaction step are Eq. (3.13), the conservation equations of the total mass of basis species, and

$$\frac{\mathbf{K1}^{(n+1)} - \mathbf{K1}^{OS}}{\Delta t} = \mathbf{SS}_{\mathbf{R,K1}}^{(n+1)} - \mathbf{SS}_{\mathbf{R,K1}}^{(n)}, \tag{3.18}$$

and

$$\frac{\mathbf{K2}^{(n+1)} - \mathbf{K2}^{OS}}{\Delta t} = \mathbf{SS}_{\mathbf{R,K2}}^{(n+1)} - \mathbf{SS}_{\mathbf{R,K2}}^{(n)}. \tag{3.19}$$

Cheng et al. [13] suggested PC is almost as efficient as OS and can provide much more accurate solutions than OS.

3.4 Complete Iteration. The CI technique can produce very accurate approximations. With CI, Eqs. (3.6) and (3.8), both linear, are solved first to get $\mathbf{T1}^{(n+1)}$ and $\mathbf{T3}^{(n+1)}$, respectively. Then Eqs. (3.1), (3.4), (3.5), and (3.7) are solved in a nonlinear iteration loop until a convergent solution is reached, where the Picard method [14] is usually employed. At each nonlinear iteration, the transport part of Eqs. (3.4), (3.5), and (3.7) is solved first to get $\mathbf{K1}^{trans}$, $\mathbf{K2}^{trans}$, and $\mathbf{T2}^{trans}$. These, along with $\mathbf{T1}^{(n+1)}$ and $\mathbf{T3}^{(n+1)}$ are then used to solve the three equations below plus the conservation equations of the total mass of basis species to yield the concentration of all species at the current iteration.

$$\Phi\left(\mathbf{C}^{(n+1)}\right) = 0, \tag{3.20}$$

$$\frac{\mathbf{K1}^{(n+1)} - \mathbf{K1}^{trans}}{\Delta t} = \mathbf{SS}_{\mathbf{R,K1}}^{(n+1)}, \tag{3.21}$$

and

$$\frac{\mathbf{K2}^{(n+1)} - \mathbf{K2}^{trans}}{\Delta t} = \mathbf{SS}_{\mathbf{R,K2}}^{(n+1)}. \tag{3.22}$$

All of the computed species concentrations, *i.e.*, **C**, are then used to update $\mathbf{K2_m}$, $\mathbf{T2_m}$, $\mathbf{SS}_{\mathbf{R,K1}}^{(n+1)}$, and $\mathbf{SS}_{\mathbf{R,K2}}^{(n+1)}$ for use in the successive nonlinear iteration. The footnote indexing `m' is to represent the mobile part of **K2** and **T2**, which appear in Eqs. (3.5) and (3.7), respecttively.

Zhang et al. [11] suggested that OS, PC, and CI be provided as options for various purposes of RT modeling. One unified way to achieve this is using PC as the backbone. By setting both $\mathbf{SS}_{\mathbf{R,K1}}^{(n)}$ and $\mathbf{SS}_{\mathbf{R,K2}}^{(n)}$ to zero, the method becomes OS. The method becomes CI by turning on nonlinear iteration that includes the transport and reaction steps, but keeps all linear equations outside the nonlinear iteration loop.

4 Summary

We compared reaction-based differential algebraic equation and primitive approaches for reactive transport computation using a hypothetical, 12-species reaction system containing equilibrium-controlled, instantaneous, and kinetic reactions. We explained why the RB-DAE approach can handle fast reactions and provide accurate solutions without the need for very small time steps. We also discussed three RT solution approaches: SNIA, SIA,

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.

and DSA. We then proposed using the predictor-corrector technique as the backbone scheme which allows for easy switching to the operator splitting or the complete iteration techniques, ensuring that the best strategy is used to meet modeling goals within constraints on computational resources, funding, and time.

The US Army Engineer Research and Development Center has used the RB-DAE approach successfully to construct a numerical simulator for biogeochemistry modeling [9], and is incorporating it into ERDC's ADaptive Hydrology/Hydraulics (ADH) Model System [15] for large-scale RT simulation using high performance computing. Both the biogeochemistry and RT simulations are used to help identify effective remediation strategies for solvent, explosive, and heavy metal contamination at current and formerly-used military sites. Typically, simulations to design remediation schemes involve meshes of more than 100,000 elements and simulation times of one to several years. Concentrations must be accurate to about one microgram per liter (part per billion) while initial concentrations may be near one gram per liter. Fast and accurate simulators are required.

References

[1] R. B. Knapp, *Spatial and temporal scales of local equilibrium in dynamic fluid-rock systems*, Geochimica et Cosmochimica Acta, 53 (1989), pp. 1955-1964.

[2] C. M. Bethke, *Geochemical and Biogeoche-mical Reaction Modeling*, 2nd edition, Cambridge University Press, 2008.

[3] J. Rubin, *Transport of reacting solutes in porous media: Relation between mathematical nature of problem formulation and chemical nature of reactions*, Water Resources Research, 19 (1983), pp. 493-502.

[4] G-T. Yeh, *Computational Subsurface Hydrology: Reactions, Transport, and Fate*, Kluwer Academic Publishers, 2000.

[5] J. Friedly and J. Rubin, *Solute transport with multiple equilibrium-controlled or kinetically controlled chemical reactions*, Water Resources Research, 28 (1992), pp. 1935– 1953.

[6] A. Chilakapati, T. Ginn, and J. E. Szecsody, *An analysis of complex reaction networks in groundwater modeling*, Water Resources Research, 34 (1998), pp. 1767-1780.

[7] Y. Fang, G-T. Yeh, and W. D. Burgos, *A general paradigm to model reaction-based biogeochemical processes in batch systems*, Water Resources Research, 39 (2003), 1083, doi:10.1029/ 2002WR0016 94.

[8] Y. Fang, S. B. Yabusaki, and G-T. Yeh, *A general simulator for reaction-based biogeoche-mical processes*, Computers & Geosciences, 32 (2006), pp. 64-72.

[9] H-P. Cheng, S. E. Howington, M. W. Farthing, C. J. McGrath, and J-R. C. Cheng, *A generic reaction-based biogeochemical simulator (RBB GCS)*, technical report, US Army Engineer Research and Development Center, 2010 (in press).

[10] T. P. Clement, *RT3D: A modular computer code for simulating reactive multi-species transport in 3-dimensional groundwater systems*, Pacific Northwest National Laboratory Technical Report, PNNL-11720, 1997.

[11] F. Zhang, G-T. Yeh, J. C. Parker, S. C. Brooks, M. N. Pace, Y-J. Kim, P. M. Jardine, and D. B. Watson, *A reaction-based paradigm to model reactive chemical transport in groundwater with general kinetic and equilibrium reactions*, Journal of Contaminant Hydrology, 92 (2007), 10-32.

[12] C. de Dieuleveult, J. Erhel, and M. Kern, *A global strategy for solving reactive transport equations*, Journal of Computational Physics, 228 (2009), 6395-6410.

[13] H-P. Cheng, M-H. Li, and J-R.C. Cheng, *An Assessment of Using the Predictor-Corrector Technique to Solve Reactive Transport Equa-tions,* International Journal for Numerical Methods in Engineering, 56 (2003), pp. 739-766.

[14] S. W. Goode and S.A. Annin, *Differential Equations and Linear Algebra*, 3rd edition, Prentice Hall, 2008.

[15] The US Army Engineer Research and Development Center ADaptive Hydrology/Hydraulics (ADH) Model System. 2009. Available from: http://www.erdc.usace.army.mil/pls/erdcpub/WWW_WELCOME.NAVIGATION_PAGE?tmp_next_page=2779688.

Copyright © by SIAM.
Unauthorized reproduction of this article is prohibited.